Article

Stability and Folding Behavior Analysis of Zinc-Finger Using Simple Models

Shan Chang [1],*, Xiong Jiao [2], Jian-Ping Hu [3], Yan Chen [1] and Xu-Hong Tian [1],*

[1] College of Informatics, South China Agricultural University, Guangzhou 510642, China;
E-Mail: ayan0426@scau.edu.cn

[2] Institute of Applied Mechanics and Biomedical Engineering, Taiyuan University of technology,
Taiyuan 030024, China; E-Mail: jiaoxiong@emails.bjut.edu.cn

[3] College of Chemistry and Life Science, Leshan Teacher's College, Leshan 614000, China;
E-Mail: lion_hjp@yahoo.com.cn

* Authors to whom correspondence should be addressed; E-Mails: schang@scau.edu.cn (S.C.);
tianxuhong@scau.edu.cn (X.-H.T.); Tel.: +86-20-85280320; Fax: +86-20-85285393.

Abstract: Zinc-fingers play crucial roles in regulating gene expression and mediating protein-protein interactions. In this article, two different proteins (Sp1f2 and FSD-1) are investigated using the Gaussian network model and anisotropy elastic network model. By using these simple coarse-grained methods, we analyze the structural stabilization and establish the unfolding pathway of the two different proteins, in good agreement with related experimental and molecular dynamics simulation data. From the analysis, it is also found that the folding process of the zinc-finger motif is predominated by several factors. Both the zinc ion and C-terminal loop affect the folding pathway of the zinc-finger motif. Knowledge about the stability and folding behavior of zinc-fingers may help in understanding the folding mechanisms of the zinc-finger motif and in designing new zinc-fingers. Meanwhile, these simple coarse-grained analyses can be used as a general and quick method for mechanistic studies of metalloproteins.

Keywords: zinc finger; folding pathway; Gaussian network model; anisotropy elastic network model; Sp1f2; FSD-1

1. Introduction

How an amino acid sequence folds into a 3D model, that is the protein folding problem, is one of the fundamental and important questions in molecular biology. During the past few years, there were substantial improvements in the field of protein folding [1–7]. Also, some new and unclear problems arose, such as chaperonin-assisted and metal-coupled protein folding. Zinc-fingers are metalloproteins. Their folding needs the binding of zinc ion (Zn(II)), which is different from spontaneous protein folding. These proteins play crucial roles in regulating gene expression and mediating protein-protein interactions [8–11]. Understanding the folding and stability of zinc-fingers is fundamental for treating severe human diseases, for example cancer and neurological disorders. Meanwhile, knowledge of zinc-finger folding can provide insight into the general mechanisms of the metal-cofactor dependent protein folding [12–14] and also help developing better strategies for the metalloprotein design [15–17].

Some experimental and theoretical works have focused on identifying which factors affect the folding and stabilization of zinc-fingers and have revealed important features [2,9,18–20]. However, because of the limited temporal and spatial resolution in experiments, the detailed folding pathway data of the zinc-finger is still lacking. Recently, based on a quantum chemical method, Dudev and Lim predicted the binding order of the four conserved ligand-binding residues [21]. Wang *et al.* used the molecular dynamics (MD) method systematically to study the folding mechanisms of zinc-fingers [2,22,23]. They made the first simulation on the coupling between the metal binding and zinc-finger folding. In their simulations, both charge transfer and metal induced protonation/deprotonation effects have been studied. They found that zinc ion not only stabilizes the component secondary and tertiary structures, but also participates in the whole folding process of zinc-finger. Although some detailed molecular dynamics simulations have been performed for the zinc-fingers, the comparison studies of folding simulation are still a challenge for metalloprotein. Many contradictions still exist in previous studies of the zinc-fingers [22–26]. For example, the two different zinc-fingers, Sp1f2 and FSD-1, have similar structures, but their folding pathways produced by MD simulations are quite different [22,23]. Even though some simulations focused on the same proteins, they obtained different results by using different MD methods. By using an optimized atomistic model to the simulation of FSD-1, Kim *et al.* suggested that the folding is initiated by the collapse of the hydrophobic core and followed by the folding of the β-hairpin and the formation of the α-helix [24]. In comparison, Lei and Duan applied the all-atom MD model with AMBER force field to FSD-1 and they proposed that the folding of the individual secondary structures is prior to the formation of native hydrophobic core [25]. There were remarkable differences between the previous MD simulations. Therefore, it is necessary to analyze different zinc-fingers with the same simulation methods and some consistent mechanisms can be revealed through the comparisons. However, the all-atom MD simulation is time-consuming and computationally expensive, so comparison work is rarely reported. Meanwhile, the involvement of metal ions in the proteins makes the simulation much more complicated. Further, some special MD simulations [22,23,27] were applied to the zinc-finger, such as the replica exchange molecular dynamics (REMD) and molecular dynamics combined with quantum chemical calculations. These special MD methods are too professional to be applied to folding comparison of zinc-fingers by general researchers. Hence, developing a fast and simple simulation method becomes valuable for metalloprotein folding. In fact, the original zinc-finger has

a different structure from the protein without zinc. Therefore, some simple model based on the structures [28–33] could provide useful information for zinc-finger folding.

In this work, we use the Gaussian network model (GNM) and anisotropy elastic network model (ANM) to analyze zinc-fingers. GNM is a schematic, coarse-grained model which is topology based and independent of sequence specificity [34–37]. GNM can provide information on conformational transition of proteins from the crystal structures and does not require the high computational cost of MD simulation. The ANM [38] is an extension of the GNM. Information about the directions of conformational transition can be obtained from analysis with ANM. Since the two classical elastic network models (ENMs) can provide the dynamic properties of proteins near an equilibrium state (usually native state), they are extensively applied to large systems that cannot be studied by all-atom MD, or applied to many proteins systematically for comparisons. However, the protein folding is usually far from the equilibrium state [3,39], so the general ENMs are not suitable for the protein folding study. Recently, some methods based on the iterative use of normal mode calculation were proposed to study protein folding/unfolding process [40,41]. These methods are the development of classical GNM, and the protein unfolding is considered to be approximated by a series of quasi-equilibrium processes corresponding to slowly increasing temperature. Su *et al*. [40] applied these methods to two classical small proteins, *i.e.*, CI2 and barnase. The iterative results obtained with the improved method were surprisingly in agreement with the MD simulation data. With a combination of classical and improved ENMs, our work may provide more comprehensive insight for the folding mechanisms of zinc-fingers.

In this article, the two different proteins, Sp1f2 [42] and FSD-1 [43], are analyzed by GNM and ANM. Our work shows how simple coarse-grained methods can investigate the mechanism of the zinc coupled folding of the zinc-finger motif. The coarse-grained methods only consider topological information and do not require the high computational cost of MD simulation. By using iterative GNM, we get the binding order of Zn(II) with the conserved ligand-binding residues and the unfolding pathway of Sp1f2 and FSD-1, which are all consistent with the experiment and MD simulation data. The previous MD simulations of the zinc-finger focused on the roles of zinc ion and the hydrophobic core, but the influence of C-terminal loop was rarely reported. In this work, we not only explain the function of zinc ion, but also discuss the effect of C-terminal loop. We find that both Zn(II) and the C-terminal loop affect the folding pathway of the different zinc-fingers. These are useful supplements for previous studies and can provide some insights into the design of zinc-fingers.

2. Systems and Methods

2.1. Protein Systems

The two proteins were chosen for our analysis in this work are shown in Figure 1. One is the second finger of the human transcription factor Sp1 (Sp1f2), which is a typical Cys2His2 type zinc-finger (PDB code: 1sp2) [42] and has 31 residues (see Figure 1A). The other is a new protein FSD-1 (PDB code: 1fsd) [43], which contains 28 residues (see Figure 1B). FSD-1 is designed based on the classical zinc-finger and replaced the binding zinc with a larger hydrophobic core. This new protein is capable of folding to a similar structure as the original zinc-finger without the help of metal ions.

Figure 1. (**A**) Ribbon representation of the main-chain fold of Sp1f2 (PDB code: 1sp2); (**B**) Ribbon representation of the main-chain fold of FSD-1 (PDB code: 1fsd).

A   B
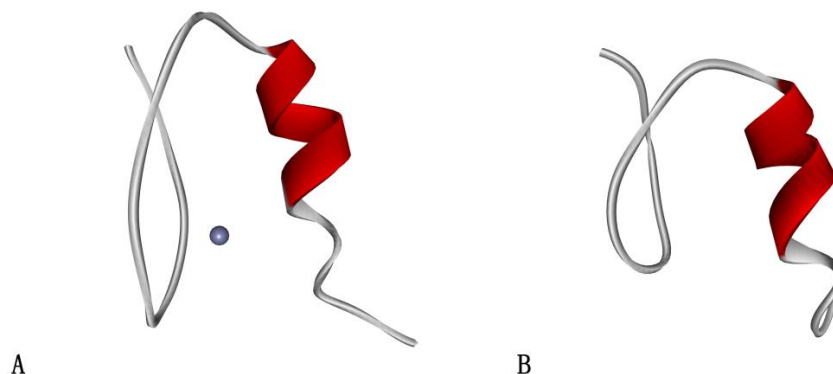
2.2. The Gaussian Network Model

In the Gaussian network model, a three-dimensional protein structure is described as an elastic network. In this work, both Cα atoms and Zn(II) are the vertices and they are connected by harmonic springs within a certain cutoff distance. The Zn(II) is encoded as 32 in the residue index of Sp1f2. The force constant is identical for all springs. Considering all contacting residues, the total energy of the network can be written as [35,44]

$$V = \frac{1}{2}\gamma\{\Delta R\}^T \Gamma\{\Delta R\}$$ (1)

where γ is the harmonic force constant; $\{\Delta R\}$ represents the N dimensional vector whose elements are the fluctuation vectors ΔR_i ($i = 1,\ldots,N$) of the individual residues, where N is the number of residues. The superscript T denotes the transpose, and Γ is the $N \times N$ symmetric matrix. The elements in the symmetric matrix are written as

$$\Gamma_{ij} = \begin{cases} -1 & if \quad i \neq j\, and \quad R_{ij} \leq r_c \\ 0 & if \quad i \neq j\, and \quad R_{ij} > r_c \\ -\sum_{i,i \neq j}\Gamma_{ij} & if \quad i = j \end{cases}$$ (2)

where R_{ij} is the separation between the ith and jth Cα atoms and r_c is the cutoff distance (7.3 Å is adopted in this work).

The mean-square fluctuation of each atom and the cross-correlation fluctuations between different atoms are in proportion to the diagonal and off-diagonal elements of the pseudo-inverse of the Γ matrix. The inverse of the matrix can be decomposed as

$$\Gamma^{-1} = U\Lambda^{-1}U^T$$ (3)

where U is an orthogonal matrix whose columns u_k ($1 \leq k \leq N$) are the eigenvectors of Γ, and Λ is a diagonal matrix of eigenvalues λ_i of Γ. The cross-correlation fluctuations between the ith and jth residues are given by

$$<\Delta R_i \cdot \Delta R_j> \quad = \quad \frac{3k_BT}{\gamma}[\Gamma^{-1}]_{ij}$$ (4)

where k_B is Boltzmann constant, T is absolute temperature. When $i = j$, the mean-square fluctuation of the ith residue can be obtained. The Debye-Waller or B-factor, which is related to the mean-square fluctuation, can be calculated with the expression

$$B_i = 8\pi^2 \langle \Delta R_i \cdot \Delta R_i \rangle / 3 \tag{5}$$

In the GNM, the cross-correlation is normalized as

$$C_{ij} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{[\langle \Delta R_i^2 \rangle \cdot \langle \Delta R_j^2 \rangle]^{1/2}} \tag{6}$$

2.3. The Iterative Unfolding Method

As the temperature of a protein is gradually increased, the native contacts between residues are expected to break in a fluctuation-dependent manner. The fluctuations in the distance between all residue pairs are calculated based on the GNM. The mean-square fluctuation in the distance vector R_{ij} between the residues i and j can be written as [45]

$$\begin{aligned}
\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle &= \left\langle \left(R_{ij} - R_{ij}^0 \right)^2 \right\rangle = \left\langle \left(\Delta R_i - \Delta R_j \right)^2 \right\rangle \\
&= \left\langle \Delta R_i \cdot \Delta R_i \right\rangle + \left\langle \Delta R_j \cdot \Delta R_j \right\rangle - 2 \left\langle \Delta R_i \cdot \Delta R_j \right\rangle \\
&= \frac{3 k_B T}{\gamma} \left(\left[\Gamma^{-1} \right]_{ii} + \left[\Gamma^{-1} \right]_{jj} - 2 \left[\Gamma^{-1} \right]_{ij} \right)
\end{aligned} \tag{7}$$

where R_{ij} and R_{ij}^0 are the instantaneous and equilibrium separation vectors between residues i and j.

The nonlinear elasticity during protein unfolding is considered through iterative normal mode calculations. As we know, the interaction between i and $i \pm 1$ residues is the covalent bond, so only the noncovalent contacts between residue pairs can be broken in the iterative process. Then, the protein unfolding process is mimicked as the following scheme:

(I)   The mean-square fluctuations of the distance in all noncovalent residue pairs are calculated based on the native structure topology with Equation 7.

(II) The noncovalent contact in the residue pair with the largest distance fluctuation is broken. Then, a new matrix Γ is obtained, which represents a new topology during protein unfolding.

(III) The mean-square fluctuations of the distance in all noncovalent residue pairs are recalculated based on the new matrix Γ using Equation 7.

(IV) The above two steps are repeated until all the noncovalent contacts are broken.

Finally, all the topologies of different conformation during protein unfolding are obtained and the unfolding pathway can be derived from the data obtained above.

2.4. The Anisotropy Elastic Network Model

The GNM model can provide the amplitudes of residue fluctuations but no information about the directions of the fluctuations. Then, the ANM model [38] is introduced, by which information about the orientation of fluctuations is elicited. In ANM, the motion mode of a protein is determined by a Hessian matrix *H*.

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ h_{N1} & h_{N2} & \cdots & h_{NN} \end{pmatrix} \tag{8}$$

The elements of *H* are submatrix with size 3×3. The *ij*th submatrix h_{ij} is

$$h_{ij} = \begin{pmatrix} \dfrac{\partial^2 V}{\partial x_i \partial x_j} & \dfrac{\partial^2 V}{\partial x_i \partial y_j} & \dfrac{\partial^2 V}{\partial x_i \partial z_j} \\ \dfrac{\partial^2 V}{\partial y_i \partial x_j} & \dfrac{\partial^2 V}{\partial y_i \partial y_j} & \dfrac{\partial^2 V}{\partial y_i \partial z_j} \\ \dfrac{\partial^2 V}{\partial z_i \partial x_j} & \dfrac{\partial^2 V}{\partial z_i \partial y_j} & \dfrac{\partial^2 V}{\partial z_i \partial z_j} \end{pmatrix} \tag{9}$$

When $i \neq j$, the analytic expression for the elements of h_{ij} is

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = \left. \frac{-\gamma(x_j - x_i)(y_j - y_i)}{R_{ij}^2} \right|_{R_{ij}=R_{ij}^0} \tag{10}$$

When $i = j$, the analytic expression for the elements of h_{ij} is

$$\frac{\partial^2 V}{\partial x_i \partial y_i} = \gamma \sum_{j \neq i} \left. \frac{(x_j - x_i)(y_j - y_i)}{R_{ij}^2} \right|_{R_{ij}=R_{ij}^0} \tag{11}$$

The meanings of γ and R are the same as that of Equation 1. x, y, and z represent the coordinates of atoms. In ANM, the cutoff distance r_c is set as 13 Å.

3. Results and Discussion

3.1. Force Constants for Inter-Residue Interactions

To evaluate the feasibility of the application of the GNM method in studying zinc-fingers, the B-factors are calculated and compared with the data from nuclear magnetic resonance (NMR). Setting $j = i$ in Equation 4, we can obtain the mean-square fluctuations of residue i, $<\Delta R_i \Delta R_i>$. According to Equation 5, the B-factors of Sp1f2 are related to the mean-square fluctuation and calculated by using $B_i = 8\pi^2 <\Delta R_i \Delta R_i>/3$. It can be seen, from the theory described above, that the only adjustable parameter in this work is γ. Then, γ can be determined by comparing the theoretically predicted mean-square fluctuations of Cα with those indicated by the experimental B-factors. The resulting $k_B T/\gamma$ value used for the zinc-fingers structure is 0.068 Å^2. Figure 2 shows the comparison between the

experimental data of Sp1f2 (light line) and the calculated B-factor of Cα atoms (dark line), in which the correlation coefficient is 0.885 for Sp1f2. The results are similar to those of recent studies for other proteins [40,46].

Figure 2. Experimental (light line) and calculated (dark line) B-factors of Sp1f2.
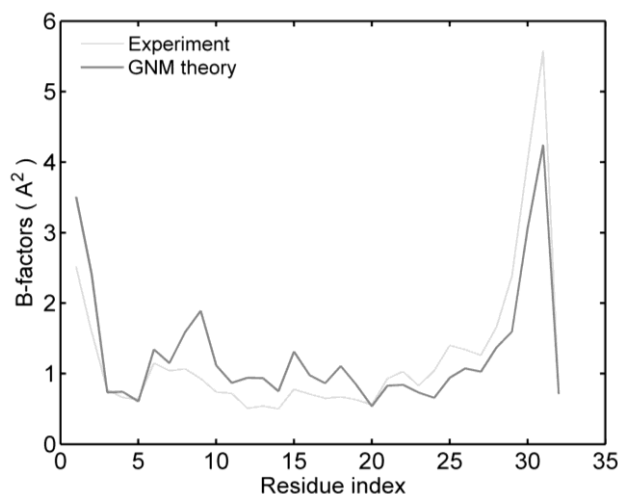
3.2. The Fast Modes of the Motions

The fast modes of the motions correspond to the geometric irregularity in the local structure. Previous studies have found that high frequency fluctuating residues are thought to be kinetically key residues and critically important for the stability of the tertiary fold [47,48]. Figure 3 shows the fastest five modes of Sp1f2. In the structure of Sp1f2, the three residues (*i.e.*, Phe3, Phe14 and Leu20) are the conserved hydrophobic ones and the conserved ligand-binding residues are Cys5, Cys10, His23 and His27 [22]. As shown in Figure 3, Phe3, Cys5, Phe14, Leu20, Zn32 are the peaks in the curves. Meanwhile, the peaks Gly11 and Lys24 are close to the residues Cys10 and His23. Our results agree well with previous studies [22] and indicate that these residues play a key role in the stability of the protein. In the classical zinc-finger motif, the hydrophobic residue Phe3 is highly conserved. Wang *et al.* gave an explanation for the conservation [22]. They indicated that Phe3 is involved in the hydrophobic interactions with the nonpolar groups of the central loop region and stabilizing the N-terminal β-hairpin should be one of the possible reasons for nature to select this hydrophobic residue. Similarly, our results can also suggest that this hydrophobic residue is a key residue for the stability of Sp1f2, which may be another support for its conservation. Figure 3 also confirms that the Zn(II) is very important for the stability of the protein. However, the fluctuation of His27 is not high in the fast modes. His27 is in the long C-terminal loop of Sp1f2, which has large and global motions. Actually, in previous studies [21,22], it was also found that His27 is the last residue coordinating to the Zn(II) in the four conserved ligand-binding residues.

Figure 3. The fastest five mode shapes of the Sp1f2. There are several peaks marked in the curve that correspond to the kinetically key residues.
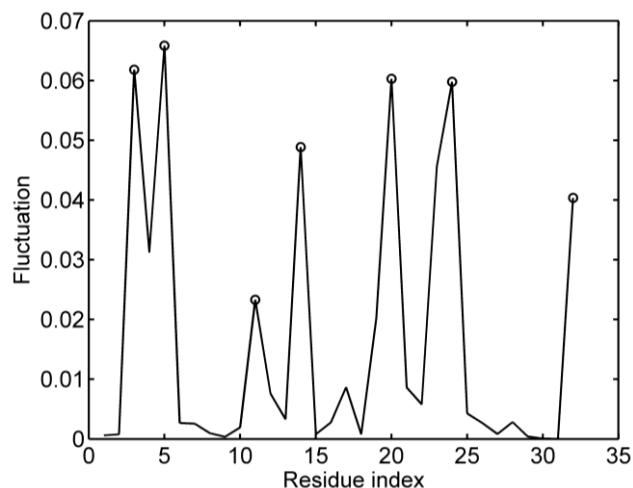

3.3. The Slow Modes of the Motions

The slow and long-wavelength collective modes usually represent functionally relevant motions of protein [48]. Figure 4 displays the slowest mode calculated by the GNM. From Figure 4, comparing the slowest modes of Sp1f2 with (holo-peptide) and without (apo-peptide) zinc binding, it can be seen that there are obvious differences between the two curves. As derived from Figure 4A, in the Sp1f2 with zinc binding, most of the residues are of low values of fluctuation. It means that these structures remain stable. However, in the apo-peptide structure (Figure 4B), the fluctuation values of residues Cys5 to Phe14 increase remarkably. The fluctuation increase of these residues implies that without zinc binding, β-hairpin becomes less stable in the apo-peptide structure than in the holo-peptide structure. It obviously confirms that Zn(II) is the key residue for the stability of β-hairpin. However, in the apo-peptide structure, the fluctuation values of residues Phe14 to His23 stay low. This can be explained by the packing of the hydrophobic Phe14, Leu20 and His23. These residues form the hydrophobic core, which is responsible for the formation and stabilization of Sp1f2 [49,50]. Meanwhile, there are some obvious increases for both structures at the end of the curve. It indicates that the C-terminal loop has global slow motions. This result can also confirm that His27 is not stable and the C-terminal loop may have an influence on the folding of Sp1f2.

*3.4. Sequences of Unfolding Events of Sp1f2 and FSD-1 Revealed by Iterative Unfolding Method*

Using the iterative unfolding method described above, the unfolding processes of Sp1f2 and FSD-1 are obtained. Based on this simple model, the results show that the two proteins have quite different folding pathways. Meanwhile, the order of the unfolding events is consistent with that of the thermal unfolding of proteins obtained with full atom MD simulation and experiments [22,23,51,52], which implies that the topology may play an important role in the two proteins folding process. This iterative unfolding method is coarse-grained and topology-based. Our results indicate that this simple model is an effectual method to explore the unfolding process of metalloprotein.

Figure 4. (**A**) The slowest mode shapes of Sp1f2 with zinc binding; (**B**) The slowest mode shapes of Sp1f2 without zinc binding.
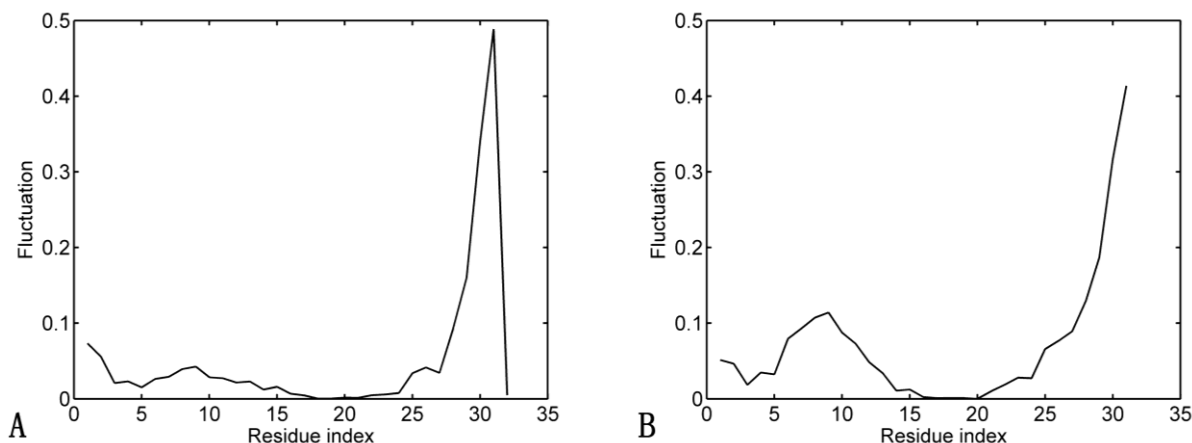
To elaborate the loss of the native contacts in the course of unfolding simulations, the contact maps of the conformation in different snapshots are constructed. Figure 5 presents the contact maps of the native structure (A), of the conformations with the loss-number-of- noncovalent-contact (LNNC) to be 20 (B), 30 (C), 48 (D), 60 (E) and 72 (F) for Sp1f2, respectively. The results reveal that there is a preferred process that shows a sequence of events for the unfolding of Sp1f2. The first unfolding event is the disappearance of the native contact in the C-terminal loop, as shown in Figure 5B. Then, the native contacts of the α-helix are partially lost, which implies that the unfolding of the C-terminal loop affect the stability of α-helix. During the partial unfolding of the α-helix, the bond between Zn(II) and His27 break first. Then, the loss of contacts between residues in β-hairpin is followed. Interestingly, as shown in Figure 5 (C, D), Zn(II) break the bond of Zn(II)-Cys5 before the bond of the Zn(II)-Cys10. This breaking order is in good agreement with the MD simulation [22] and experimental observation [51,52]. In addition, the folding cooperativity is considered as an important behavior of protein folding dynamics [53]. Previous studies [22] found that the zinc binding and the hydrophobic core formation are quite cooperative. In our model, the unfolding pathway is sequential and it is difficult to observe the cooperativity directly in the unfolding process. In fact, these highly cooperative behaviors take place in the near-neighbor steps for this iterative unfolding model. As shown in Figure 5B, when Cys5, Cys10 and His23 bind to Zn(II), the hydrophobic core is stable and integrated. After the breaking up of the Zn(II)-His23 bond, both of the hydrophobic core and the bonds between Zn(II) and two cysteines begin to unfold (see Figure 5C). When the bond of Zn(II)-Cys5 breaks, the native contacts of His23 are completely lost and the hydrophobic core is partly unfolding (see Figure 5D). These two behaviors happen in the near-neighbor phases and they may show cooperation to some extent. At the end of unfolding (see Figure 5 (E, F)), the native contacts of β-hairpin disappear and only α-helix is partially retained. This unfolding pathway is excitingly consistent with the pathway of MD simulation [22]. The results indicate that the unfolding pathway is mainly determined by its native topology and the iterative unfolding method can be used to reasonably describe the zinc binding process.

Figure 5. The contact maps of the native conformation (**A**), of the conformations with the LNNC to be 20 (**B**), 30 (**C**), 48 (**D**), 60 (**E**) and 72 (**F**) for Sp1f2, respectively. Each native contact is marked by the symbol * in the maps.
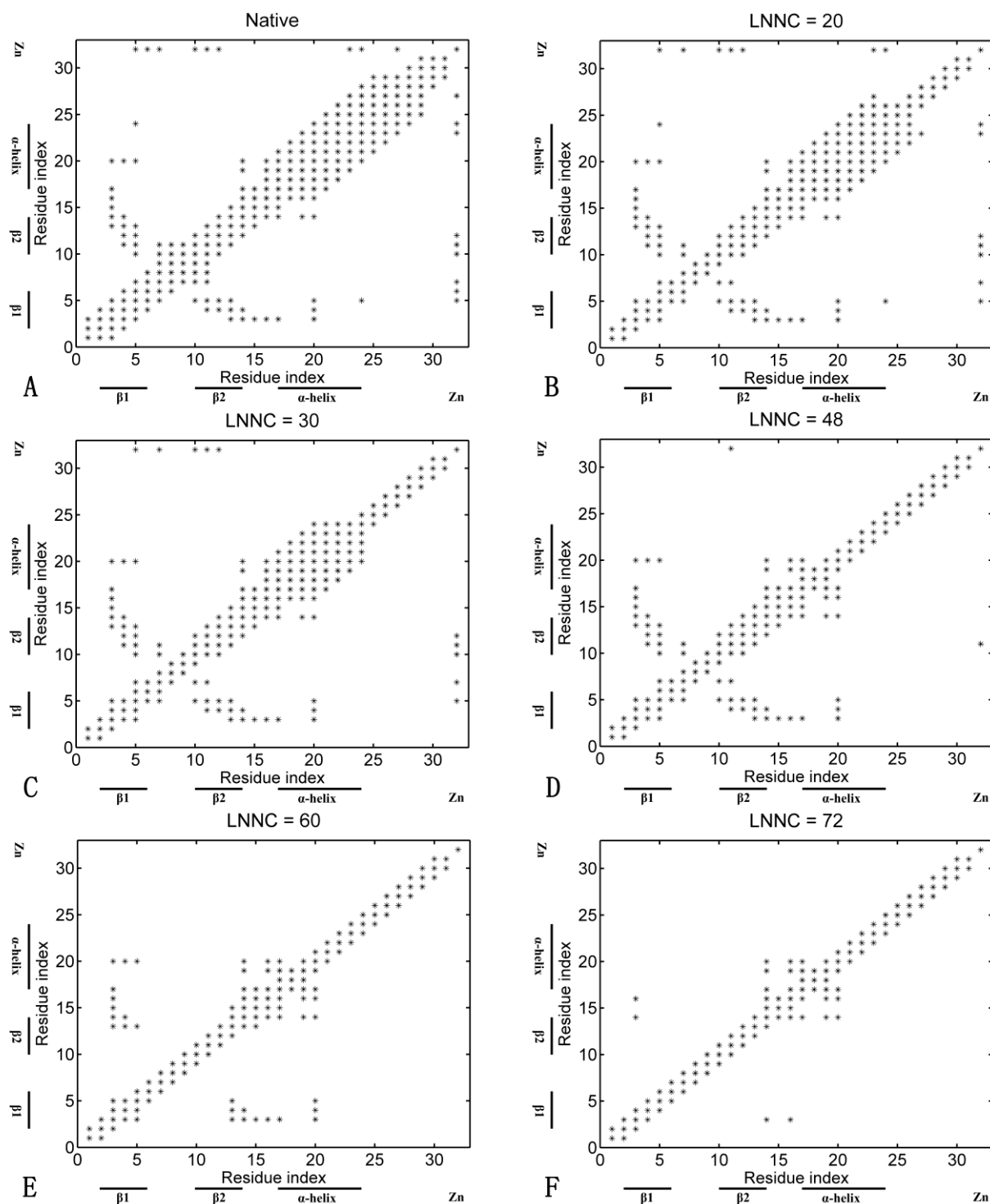
The contact maps of the conformation in different snapshots during unfolding for FSD-1 are shown in Figure 6. This figure presents the contact maps of the native structure (Figure 6A), of the conformations with the LNNC to be 20 (Figure 6B), 30 (Figure 6C), 50 (Figure 6D) for FSD-1, respectively. The unfolding process of FSD-1 can be obtained from the change of the contact map, which is quite different from that of Sp1f2. As shown in Figure 6B, the unfolding is initiated by the

partial disruption of the native contact between β-hairpin. After the loss of β-hairpin, the native contact of α-helix begins to break, which is accompanied with the disappearance of the hydrophobic core (see Figure 6 (C, D)). The previous MD simulations [23] found that in the folding process of FSD-1, the α-helix is constructed before the formation of β-hairpin. Our unfolding results also confirm this folding pathway and indicate that the α-helix is more stable compared with the β-hairpin. The previous studies [23] also found that the middle helical turn of the α-helix (R19, D20, F21 and I22) was more stable than the first helical turn (E15, K16, E17 and L18) at high temperatures. Interestingly, as shown in Figure 6 (C, D), the first helical turn is unfolding before that of the middle helical turn.

Figure 6. The contact maps of the native conformation (**A**), of the conformations with the LNNC to be 20 (**B**), 30 (**C**) and 50 (**D**) for FSD-1, respectively. Each native contact is marked by the symbol * in the maps.
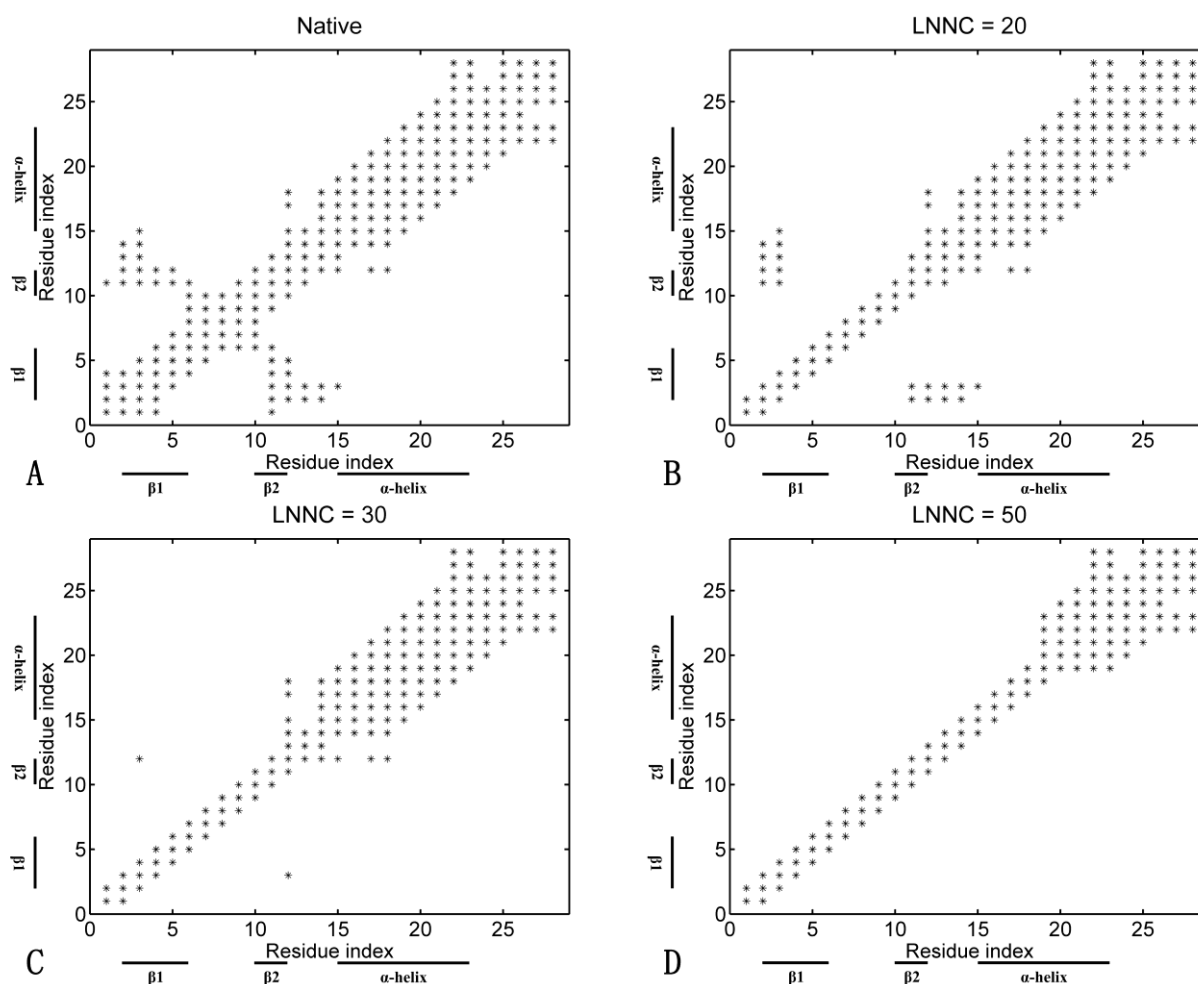
In addition, the contact maps of unfolding snapshots for Sp1f2 without zinc binding are shown in Figure 7. Although this process is not really existing, the simulation can also help us to understand some crucial factors for the folding. As shown in Figure 7, the unfolding pathway of apo-peptide is different with those of the above two proteins. Without zinc binding, the structure of the β-hairpin is unstable. Meanwhile, the long C-terminal loop also affects the stability of α-helix, so the β-hairpin and α-helix are both unfolding in the early stage of the simulation process. As shown in Figure 7 (B–E), the native contacts of the β-hairpin disappeared faster than that with zinc binding. In contrast, the α-helix

unfolded slower than that with zinc binding (see Figure 7 (B, C)). These results show that Zn(II) stabilizes the whole structure of Sp1f2. Without zinc binding, both β-hairpin and α-helix are unstable, so it is impossible for Sp1f2 to complete the correct folding.

Figure 7. The contact maps of the native conformation (**A**), of the conformations with the LNNC to be 20 (**B**), 30 (**C**), 48 (**D**), 60 (**E**) and 72 (**F**) for Sp1f2 without zinc binding, respectively. Each native contact is marked by the symbol * in the maps.
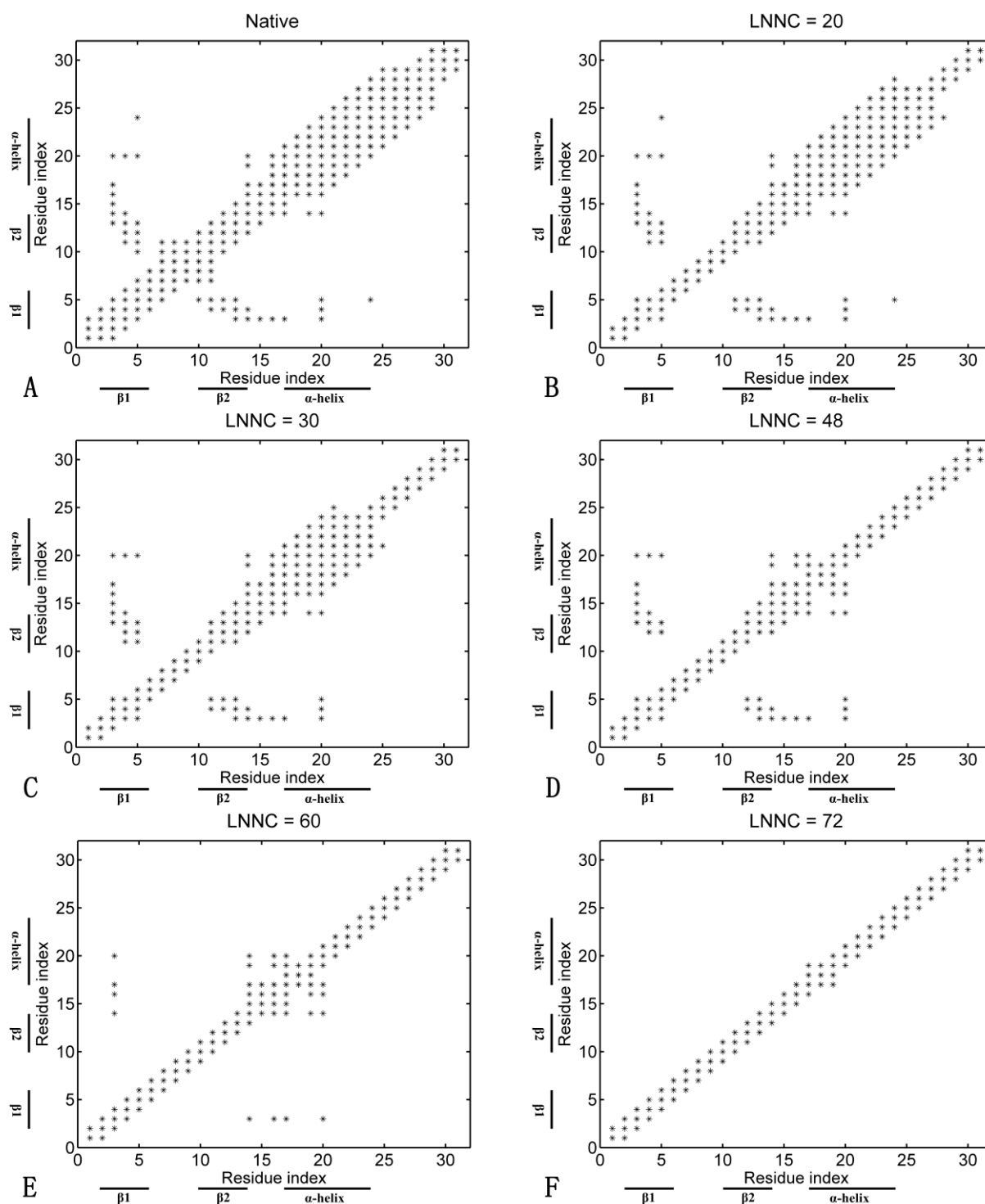
3.5. Cross-Correlation Analysis during Protein Unfolding

Similar to the analysis of Su *et al*. [40,54], we explore the change of correlation between the fluctuations of residues during the unfolding process of the two proteins. The cross-correlations between the fluctuations of residues are calculated with Equation 6. The cross-correlation value ranges from −1 to 1. The positive values represent that the motions of residues are in the same direction, and the negative values represent that they move in the opposite direction. The higher the absolute cross-correlation value is, the more the two residues are correlated (or anti-correlated). On the other hand, the cross-correlation value $C_{ij} = 0$ means that the motions of residues are completely not correlated.

Figure 8 presents the cross-correlation maps of Sp1f2. As shown in Figure 8A, along the diagonal of the map, there are some light blocks with positive correlations, which correspond to the secondary structures of α-helix and β-sheets. The figure also shows negative correlations between the β-hairpin and C-terminal loop. Meanwhile, the β-hairpin has higher positive correlations with Zn(II) than those of α-helix. As shown in Figure 8B, with connection of Zn(II), the β-hairpin and α-helix become a whole part and have positive correlations between them. Then, as shown in Figure 8 (C, D), with influence of C-terminal loop, the α-helix is partially unfolding and the structure of the protein seems to be divided into two parts. The β-hairpin and α-helix cooperate together and they have negative correlations with C-terminal loop. At the end of unfolding (see Figure 8 (E, F)), the β-hairpin is unfolding and only partial α-helix is kept.

The cross-correlation maps during the unfolding process of FSD-1 are shown in Figure 9. It is easy to identify the secondary structures and native contacts with positive correlations from the correlation map of the native structure shown in Figure 9A. As shown in Figure 9 (B, C), when the native contacts in the β-hairpin are lost, the positive correlations between the β-hairpin are reduced. Finally, the structure of the protein seems to be divided into two parts fluctuating in opposite directions (see Figure 9D). From Figure 9, we can see that in the unfolding process of FSD-1, the negative correlations between the β-hairpin and α-helix are always maintained. These phenomena show that the β-hairpin and α-helix have weak cooperativity, which is different from that of Sp1f2. These results are also consistent with previous MD simulations [22,23].

Similar to Figure 7, the cross-correlation maps of Sp1f2 without zinc binding are also presented in Figure 10. Without zinc binding, the correlation between β-hairpin and α-helix is reduced. As shown in Figure 10 (B, C), the cross-correlation maps of apo-peptide are similar to those of FSD-1. However, because of the large fluctuation of C-terminal loop, the structure of the protein is also divided into two parts (see Figure 10 (D, E)). Then, the following cross-correlation maps are similar to those of Sp1f2 with zinc binding. The absence of zinc ion and the long C-terminal loop put the dual influences on the structure, so the cross-correlation maps of Sp1f2 without zinc binding shows the combined behaviors between those of FSD-1 and Sp1f2 with zinc binding.

Figure 8. The cross-correlation maps calculated using all modes for native conformation (**A**) and several conformations with LNNC to be 20 (**B**), 30 (**C**), 48 (**D**), 60 (**E**) and 72 (**F**) during the unfolding process of Sp1f2. As shown in the bar on the right, the dark regions in the figure indicate negative correlation and the light regions present positive correlation. Both the x and y axes of the maps are residue indices.
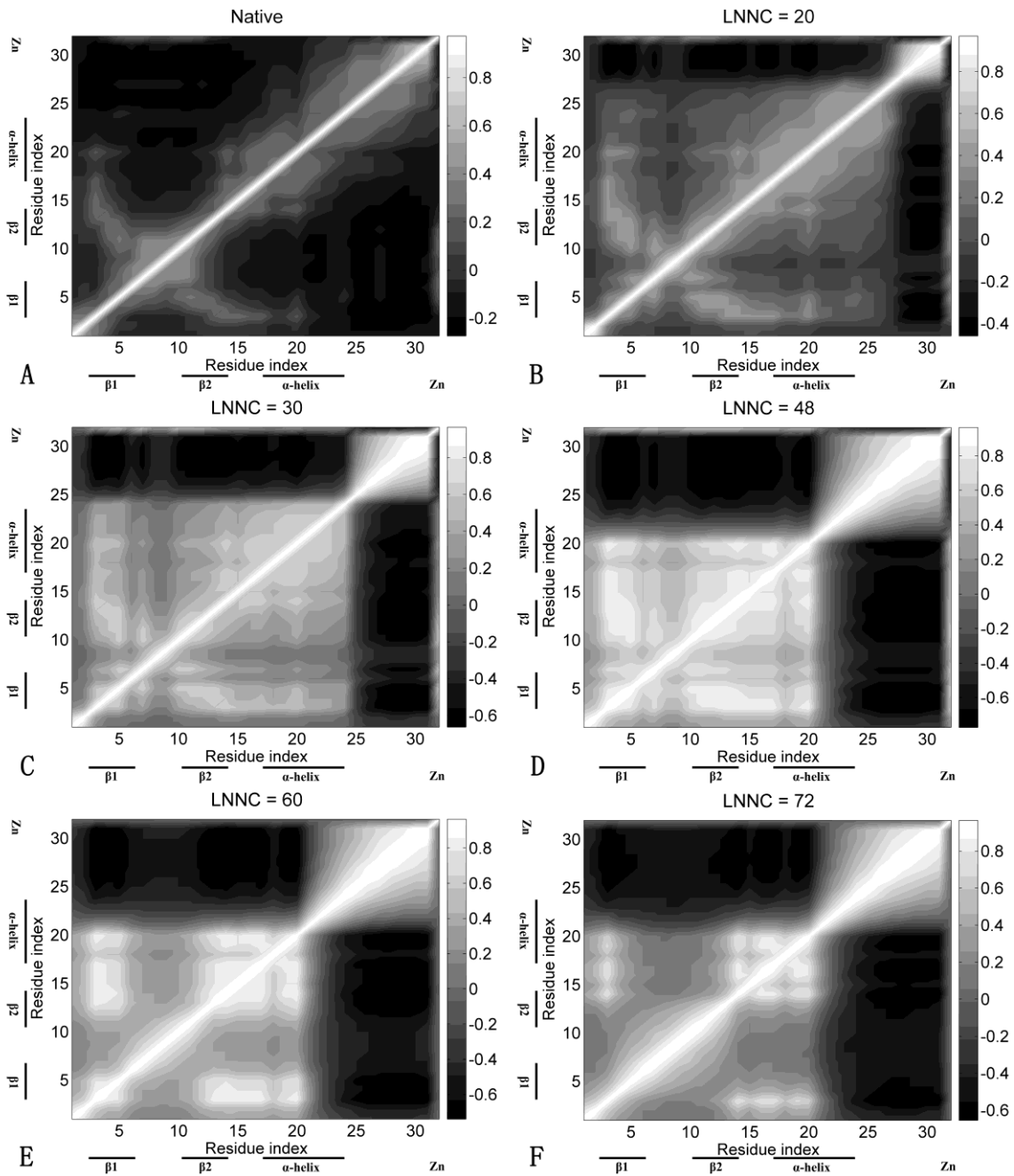
Figure 9. The cross-correlation maps for native conformation (**A**) and several conformations with LNNC to be 20 (**B**), 30 (**C**) and 50 (**D**) during the unfolding process of FSD-1. As shown in the bar on the right, the dark regions in the figure indicate negative correlation and the light regions present positive correlation. Both the x and y axes of the maps are residue indices.
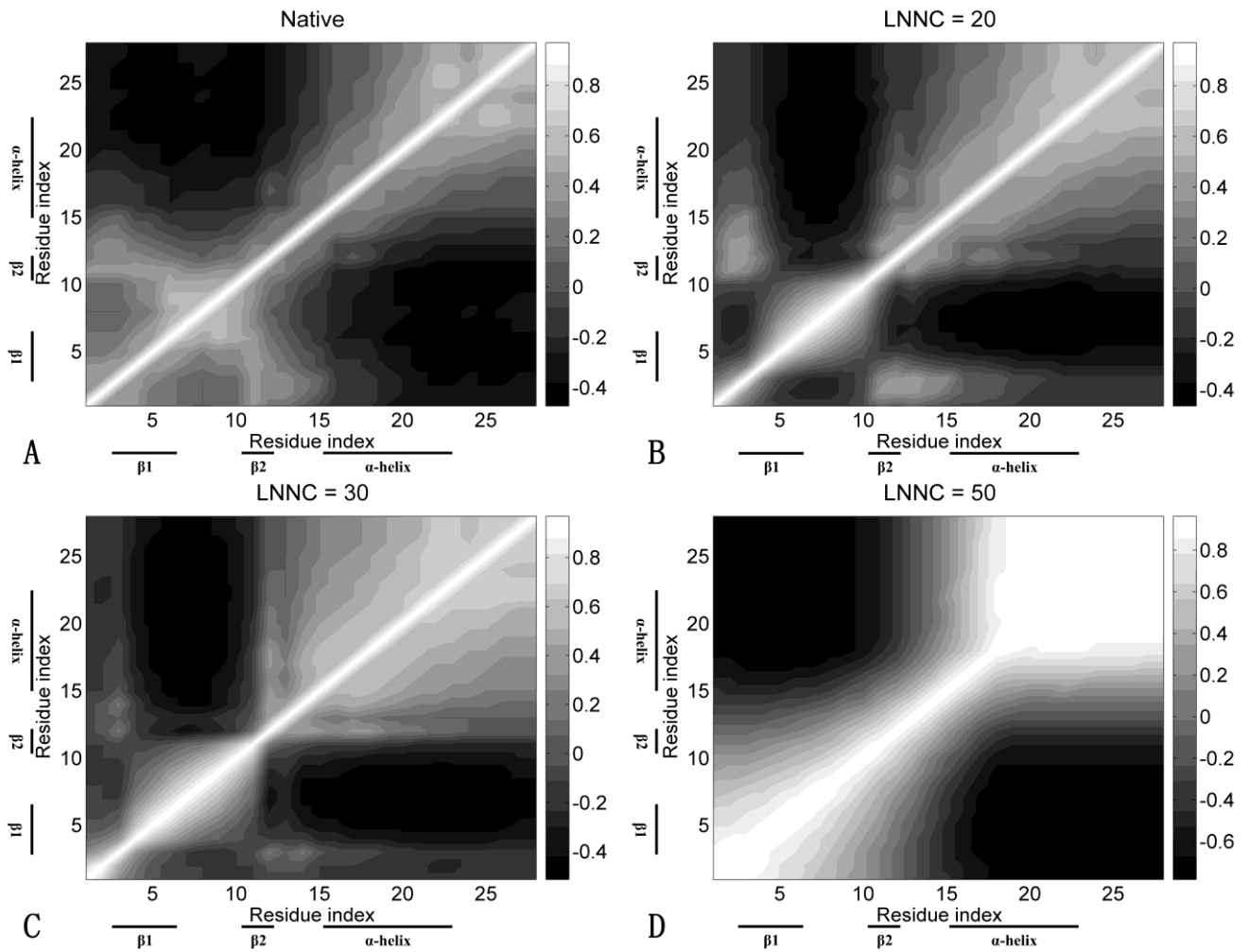
Figure 10. The cross-correlation maps for native conformation (**A**) and several conformations with LNNC to be 20 (**B**), 30 (**C**), 48 (**D**), 60 (**E**) and 72 (**F**) during the unfolding process of Sp1f2 without zinc binding. As shown in the bar on the right, the dark regions in the figure indicate negative correlation and the light regions present positive correlation. Both the x and y axes of the maps are residue indices.
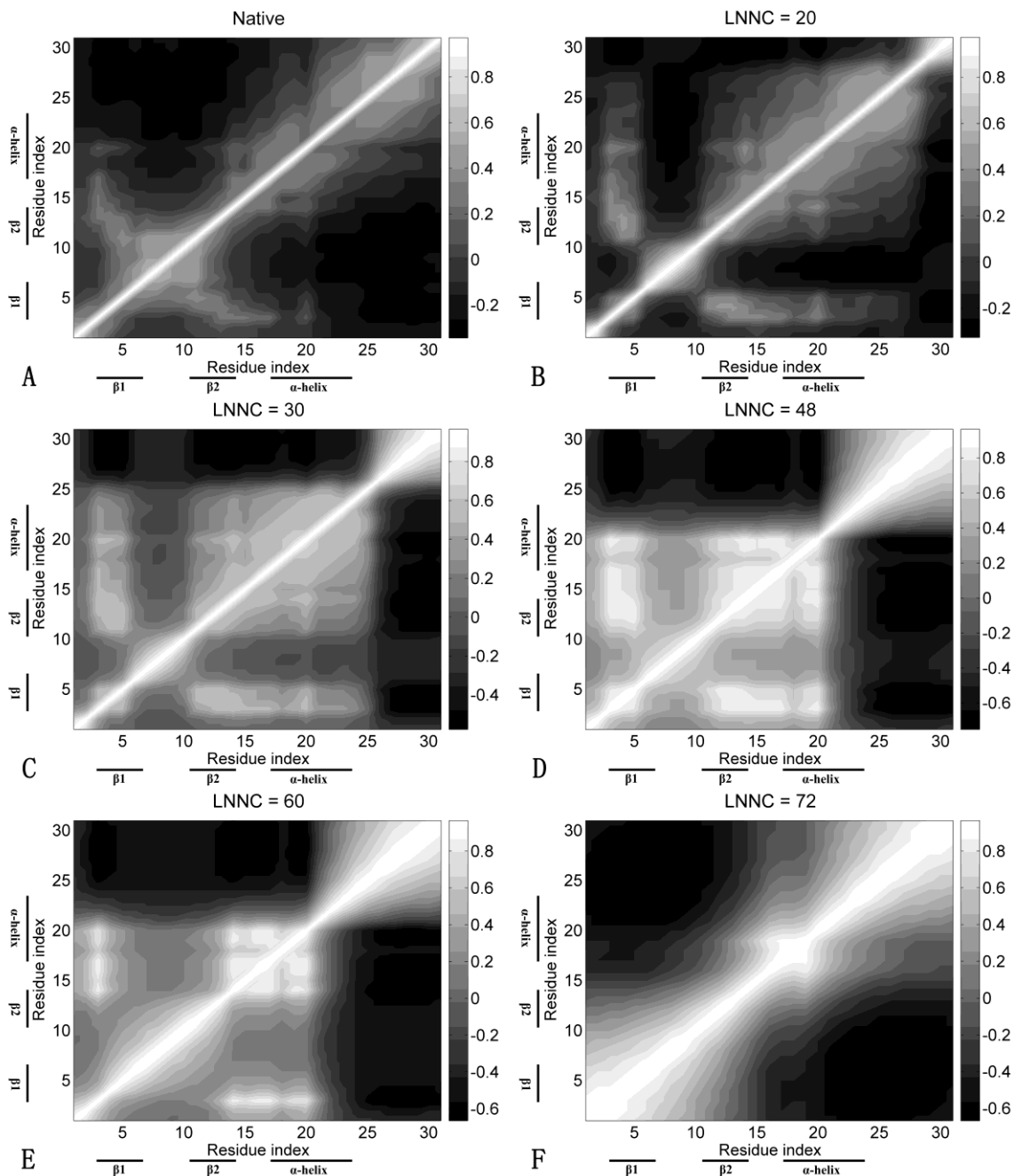

3.6. The ANM Analysis of Sp1f2 and FSD-1

GNM can only provide the magnitude of displacement of atoms from their equilibrium positions for large-scale motions. To ascertain the direction of motion, ANM is applied to the structures. For the Sp1f2 structure, as seen in Figure 11A, the first slowest mode of β-hairpin and α-helix corresponds to the swing motion. However, the C-terminal loop shows an opposite direction with the other parts of

the structure. The motive magnitude of the C-terminal loop is much larger than those of the other parts. It also confirms that the C-terminal loop has large fluctuation, which will affect the stability of the α-helix. As seen from Figure 11B, the first slowest motion of FSD-1 is obviously smaller than that of Sp1f2. The most flexible regions for FSD-1 are the N-terminal loop. The fluctuation of N-terminal loop will affect the stability of the β-hairpin. Compared with Sp1f2, it may explain why the α-helix is more stable compared with the β-hairpin in the FSD-1.

Figure 11. The first slowest motive mode sketch maps of the Sp1f2 (**A**) and FSD-1 (**B**). The first slow motive modes are shown with the cone model. The length of cone is correlative with the motive magnitude and the motive direction is depicted with the orientation of cone.

A   B
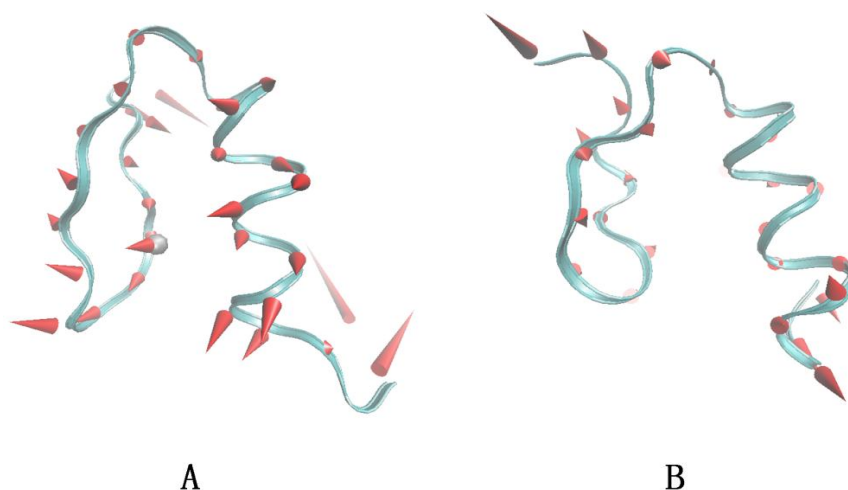
3.7. The Analysis of Folding Pathways

From the above simulations, it is found that the influence of Zn(II) on the folding of zinc-fingers lies in two points. First, Zn(II) preferentially binds to the two conserved cysteine residues and helps the formation of the β-hairpin in the early stage of folding. Second, Zn(II) makes the connection between the β-hairpin and α-helix, which increases the cooperativity of the two parts and stabilizes the whole structure of Sp1f2. On the other hand, the C-terminal loop has a negative effect for the stability of zinc-finger. In the Sp1f2, the long C-terminal loop is near to the α-helix and has large fluctuation, which results in an unfavorable effect on the formation of the α-helix. Both the Zn(II) and C-terminal loop predominate the folding pathway of the zinc-finger motif. In the folding of Sp1f2, Zn(II) accelerates the formation of the β-hairpin. Meanwhile, Zn(II) connects the β-hairpin and α-helix, and it helps the α-helix to resist the negative effect of the long C-terminal loop. Therefore, as shown in Figure 5, the folding of Sp1f2 initiates with a hydrophobic collapse. Then, with connection of Zn(II), the β-hairpin is fully formed and the α-helix is partially folded. Finally, the folding finishes with full formation of the α-helix and C-terminal loop. However, in the folding of FSD-1, the short C-terminal loop favors the stabilization of α-helix. Meanwhile, without Zn(II), the formation of β-hairpin is decelerated. Then, the folding of FSD-1 initiates with the formation of the α-helix, and it finishes with the formation of the β-hairpin (see Figure 6).

4. Conclusions

In this work, simple coarse-grained methods are proposed to investigate the mechanism of the zinc-finger motif. By using the Gaussian network model and anisotropy elastic network model, we analyze the binding order of Zn(II) with the conserved ligand-binding residues and the unfolding pathway of Sp1f2 and FSD-1, which are consistent with related experimental and MD simulation data. From the analysis, it is found that Zn(II) has two positive influences on the folding of zinc-fingers. On the contrary, the C-terminal loop has a negative effect on the stability of zinc-fingers. Both Zn(II) and the C-terminal loop affect the folding pathway of the different proteins. These folding and stability factors of the zinc-finger revealed in this work can provide insight into the folding mechanisms of the zinc-finger motif and help in the design of zinc-fingers. Meanwhile, these simple coarse-grained methods may be helpful for understanding the mechanisms of metal-coupled and chaperonin-assisted protein folding.

Acknowledgements

References

1. Zhang, Y. Protein structure prediction: When is it useful? *Curr. Opin. Struct. Biol.* **2009**, *19*, 145–155.
2. Zhang, J.; Li, W.F.; Wang, J.; Qin, M.; Wu, L.; Yan, Z.Q.; Xu, W.X.; Zuo, G.H.; Wang, W. Protein folding simulations: from coarse-grained model to all-atom model. *IUBMB Life* **2009**, *61*, 627–643.
3. Chen, Y.W.; Ding, F.; Nie, H.F.; Serohijos, A.W.; Sharma, S.; Wilcox, K.C.; Yin, S.Y.; Dokholyan, N.V. Protein folding: then and now. *Arch. Biochem. Biophys.* **2008**, *469*, 4–19.
4. Gruebele, M. Protein folding: the free energy surface. *Curr. Opin. Struct. Biol.* **2002**, *12*, 161–168.
5. Alexandrescu, A.T. Amyloid accomplices and enforcers. *Protein Sci.* **2005**, *14*, 1–12.
6. Xiao, Y.; Chen, C.J.; He, Y. Folding Mechanism of Beta-Hairpin Trpzip2: Heterogeneity, Transition State and Folding Pathways. *Int. J. Mol. Sci.* **2009**, *10*, 2838–2848.
7. Yang, W.Y.; Gruebele, M. Folding at the speed limit. *Nature* **2003**, *423*, 193–197.
8. Anzellotti, A.I.; Farrell, N.P. Zinc metalloproteins as medicinal targets. *Chem. Soc. Rev.* **2008**, *37*, 1629–1651.
9. Frankel, A.D.; Berg, J.M.; Pabo, C.O. Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4841–4845.
10. Laity, J.H.; Lee, B.M.; Wright, P.E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **2001**, *11*, 39–46.
11. Leon, O.; Roth, M. Zinc fingers: DNA binding and protein-protein interactions. *Biol. Res.* **2000**, *33*, 21–30.

12. Wittung-Stafshede, P. Role of cofactors in folding of the blue-copper protein azurin. *Inorg. Chem.* **2004**, *43*, 7926–7933.

13. Wilson, C.J.; Apiyo, D.; Wittung-Stafshede, P. Role of cofactors in metalloprotein folding. *Q. Rev. Biophys.* **2004**, *37*, 285–314.

14. Li, W.F.; Zhang, J.; Su, Y.; Wang, J.; Qin, M.; Wang, W. Effects of zinc binding on the conformational distribution of the amyloid-beta peptide based on molecular dynamics simulations. *J. Phys. Chem. B* **2007**, *111*, 13814–13821.

15. Ghosh, D.; Pecoraro, V.L. Probing metal-protein interactions using a *de novo* design approach. *Curr. Opin. Chem. Biol.* **2005**, *9*, 97–103.

16. Papworth, M.; Kolasinska, P.; Minczuk, M. Designer zinc-finger proteins and their applications. *Gene* **2006**, *366*, 27–38.

17. Ghosh, D.; Pecoraro, V.L. Understanding metalloprotein folding using a *de novo* design strategy. *Inorg. Chem.* **2004**, *43*, 7902–7915.

18. Bombarda, E.; Roques, B.P.; Mely, Y.; Grell, E. Mechanism of zinc coordination by point-mutated structures of the distal CCHC binding motif of the HIV-1NCp7 proteins. *Biochemistry* **2005**, *44*, 7315–7325.

19. Nomura, A.; Sugiura, Y. Contribution of individual zinc ligands to metal binding and peptide folding of zinc finger peptides. *Inorg. Chem.* **2002**, *41*, 3693–3698.

20. Blasie, C.A.; Berg, J.M. Structure-based thermodynamic analysis of a coupled metal binding-protein folding reaction involving a zinc finger peptide. *Biochemistry* **2002**, *41*, 15068–15073.

21. Dudev, T.; Lim, C. All-electron calculations of the nucleation structures in metal-induced zinc-finger folding: Role of the peptide backbone. *J. Am. Chem. Soc.* **2007**, *129*, 12497–12504.

22. Li, W.F.; Zhang, J.; Wang, J.; Wang, W. Metal-coupled folding of Cys(2)His(2) zinc-finger. *J. Am. Chem. Soc.* **2008**, *130*, 892–900.

23. Li, W.F.; Zhang, J.; Wang, W. Understanding the folding and stability of a zinc finger-based full sequence design protein with replica exchange molecular dynamics simulations. *Proteins* **2007**, *67*, 338–349.

24. Kim, S.Y.; Lee, J.; Lee, J. Folding simulations of small proteins. *Biophys. Chem.* **2005**, *115*, 195–200.

25. Lei, H.X.; Duan, Y. The role of plastic beta-hairpin and weak hydrophobic core in the stability and unfolding of a full sequence design protein. *J. Chem. Phys.* **2004**, *121*, 12104–12111.

26. Jang, S.; Shin, S.; Pak, Y. Molecular dynamics study of peptides in implicit water: ab initio folding of beta-hairpin, beta-sheet, and beta beta alpha-motif. *J. Am. Chem. Soc.* **2002**, *124*, 4976–4977.

27. Feng, J.W.A.; Kao, J.; Marshall, G.R. A Second Look at Mini-Protein Stability: Analysis of FSD-1 Using Circular Dichroism, Differential Scanning Calorimetry, and Simulations. *Biophys. J.* **2009**, *97*, 2803–2810.

28. Hills, R.D.; Brooks, C.L. Insights from Coarse-Grained Go Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889–905.

29. Clementi, C. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10–15.

30. Chang, S.; Jiao, X.; Gong, X.Q.; Li, C.H.; Chen, W.Z.; Wang, C.X. Evolving model of amino acid networks. *Phys. Rev. E* **2008**, *77*, 061920.

31. Jiao, X.; Chang, S.; Li, C.H.; Chen, W.Z.; Wang, C.X. Construction and application of the weighted amino acid network based on energy. *Phys. Rev. E* **2007**, *75*, 051903.

32. Zhang, M.; Chen, C.J.; He, Y.; Xiao, Y. Improvement on a simplified model for protein folding simulation. *Phys. Rev. E* **2005**, *72*, 051919.

33. Bahar, I.; Rader, A.J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.

34. Yang, L.W.; Liu, X.; Jursa, C.J.; Holliman, M.; Rader, A.; Karimi, H.A.; Bahar, I. iGNM: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics* **2005**, *21*, 2978–2987.

35. Haliloglu, T.; Bahar, I.; Erman, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090–3093.

36. Bahar, I.; Lezon, T.R.; Bakan, A.; Shrivastava, I.H. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem. Rev.* **2010**, *110*, 1463–1497.

37. Chennubhotla, C.; Yang, Z.; Bahar, I. Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. Biosyst.* **2008**, *4*, 287–292.

38. Atilgan, A.R.; Durell, S.R.; Jernigan, R.L.; Demirel, M.C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515.

39. Micheletti, C.; Lattanzi, G.; Maritan, A. Elastic properties of proteins: insight on the folding process and evolutionary selection of native structures. *J. Mol. Biol.* **2002**, *321*, 909–921.

40. Su, J.G.; Li, C.H.; Hao, R.; Chen, W.Z.; Wang, C.X. Protein unfolding behavior studied by elastic network model. *Biophys. J.* **2008**, *94*, 4586–4596.

41. Sulkowska, J.I.; Kloczkowski, A.; Sen, T.Z.; Cieplak, M.; Jernigan, R.L. Predicting the order in which contacts are broken during single molecule protein stretching experiments. *Proteins* **2008**, *71*, 45–60.

42. Narayan, V.A.; Kriwacki, R.W.; Caradonna, J.P. Structures of zinc finger domains from transcription factor Sp1. Insights into sequence-specific protein-DNA recognition. *J. Biol. Chem.* **1997**, *272*, 7801–7809.

43. Dahiyat, B.I.; Mayo, S.L. *De novo* protein design: fully automated sequence selection. *Science* **1997**, *278*, 82–87.

44. Erman, B. The Gaussian network model: precise predictions of residue fluctuations and application to binding problems. *Biophys. J.* **2006**, *91*, 3589–3599.

45. Kloczkowski, A.; Mark, J.E.; Erman, B. Chain dimensions and fluctuations in random elastomeric networks. Phantom Gaussian networks in the undeformed state. *Macromolecules* **1989**, *22*, 1423–1432.

46. Bahar, I.; Atilgan, A.R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **1997**, *2*, 173–181.

47. Haliloglu, T.; Keskin, O.; Ma, B.Y.; Nussinov, R. How similar are protein folding and protein binding nuclei? Examination of vibrational motions of energy hot spots and conserved residues. *Biophys. J.* **2005**, *88*, 1552–1559.

48. Bahar, I.; Atilgan, A.R.; Demirel, M.C.; Erman, B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* **1998**, *80*, 2733.

49. Jasanoff, A.; Weiss, M.A. Aromatic histidine interactions in the zinc finger motif: structural inequivalence of phenylalanine and tyrosine in the hydrophobic core. *Biochemistry* **1993**, *32*, 1423–1432.

50. Jasanoff, A.; Kochoyan, M.; Fraenkel, E.; Lee, J.P.; Weiss, M.A. Aromatic aromatic interactions in the zinc finger motif: Analysis of the 2-dimensional nuclear magnetic resonance structure of a mutant domain. *J. Mol. Biol.* **1992**, *225*, 1035–1047.

51. Miura, T.; Satoh, T.; Takeuchi, H. Role of metal-ligand coordination in the folding pathway of zinc finger peptides. *Biochim. Biophys. Acta. Protein. Struct. Mol. Enzymol.* **1998**, *1384*, 171–179.

52. Shi, Y.G.; Beger, R.D.; Berg, J.M. Metal-binding properties of single amino-acid deletion mutants of zinc finger peptides: studies using cobalt(II) as a spectroscopic probe. *Biophys. J.* **1993**, *64*, 749–753.

53. Wu, L.; Li, W.F.; Liu, F.; Zhang, J.; Wang, J.; Wang, W. Understanding protein folding cooperativity based on topological consideration. *J. Chem. Phys.* **2009**, *131*, 065105.

54. Su, J.G.; Jiao, X.; Sun, T.G.; Li, C.H.; Chen, W.Z.; Wang, C.X. Analysis of domain movements in glutamine-binding protein with simple models. *Biophys. J.* **2007**, *92*, 1326–1335.